DOCUMENT RESUME

ED 095 362 95 CE 001 926

AUTHOR Johansson, Charles B.
TITLE Technical Aspects: Problems of Scale Development, Norms, Item Differences by Sex, and the Rate of Change in Occupational Group Characteristics (Revised February 6, 1974).
INSTITUTION Aries Corp., Minneapolis, Minn.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 13 Dec 73
CONTRACT OEC-0-72-5240
NOTE 60p.; For related documents see CE 001 927 through 935; Draft

EDRS PRICE MF-$0.75 HC-$3.15 PLUS POSTAGE
DESCRIPTORS Career Choice; Equal Opportunities (Jobs); *Interest Research; *Interest Scales; *Interest Tests; Occupational Choice; Sex Differences; *Sex Discrimination; Sex Stereotypes; *Vocational Interests
IDENTIFIERS Sex

ABSTRACT

The scope of the paper is to review the two major interest inventories, exploring the nuances and complexities of the technical aspects in their development, their item sampling, norming, scoring, reporting of results, and changing patterns of interests in relation to the differential treatment of sexes; and to suggest guidelines to eliminate or alleviate any potential sex biasing factors. Using the operational definition of sex bias as any factor that may influence a person to limit career opportunities solely on the basis of gender, these areas are explored and the following recommendations, among others, made: Interest inventories should not have separate forms of the instrument for males and females, and the same set of items should be used for both, with care that an item is not inherently more applicable to one gender than the other; in the development of externally based scales, every effort should be made to collect adequate samples of males and females for each occupation represented on the inventory's reporting of results; and caution should be exercised against differentially including items that represent non-valid sex differences. (AJ)

DRAFT

TECHNICAL ASPECTS: PROBLEMS OF SCALE DEVELOPMENT, NORMS, ITEM DIFFERENCES BY SEX, AND THE RATE OF CHANGE IN OCCUPATIONAL GROUP CHARACTERISTICS

Dr. Charles B. Johansson
National Computer Systems
4401 West 76th Street
Minneapolis, Minnesota 55435
(612) 920-3670

December 13, 1973

(Revised February 6, 1974)

National Institute of Education

Contract No. OEC-0-72-5240

This paper was funded under a contract to the ARIES Corporation (#OE-C-72-5240) from the National Institute of Education of the U.S. Department of Health, Education, and Welfare. The views expressed do not necessarily reflect those of the ARIES Corporation, the National Institute of Education, or any agency of the U.S. Government.

TECHNICAL ASPECTS: PROBLEMS OF SCALE DEVELOPMENT, NORMS, ITEM DIFFERENCES BY SEX, AND THE RATE OF CHANGE IN OCCUPATIONAL GROUP CHARACTERISTICS[1]

Dr. Charles B. Johansson

National Computer Systems

INTRODUCTION

As part of the larger concern over various types of discrimination and bias in the occupational world, the impact that interest inventories have in guiding people into careers and vocations is being studied. For some people, the choice of a vocation may be a moot question; for example, the classic case of the offspring who has known from childhood that he or she will become heir to the parent's business may have little choice in a career decision; or the person who is in the right place at the right time may have opportunities open up that no interest inventory could predict or help in the decision of a lifetime career.

But for the majority of people, their choice of a decision as to what field of work they should enter is perhaps as important a decision as whom they should marry and both have considerable impact on their lives. Professionally trained counselors continually help people who are dissatisfied with their current jobs and desire a change to something they will find more personally rewarding. Students who are graduating from high school are typically indecisive about what curriculum they should enter in college or in a business/technical school and thus seek guidance. In addition, more and more women are re-entering the occupational world and college and seek professional help in their pursuit of career opportunities.

To help the counselors in their role of providing career-plan-

ning assistance to their clients, interest inventories have been relied upon to provide data on client's vocational interest preferences and to help increase the probability that a chosen career or job will be a satisfying one.

Typically, the major interest inventories have separated responses of males and females during the developmental stages and typically report different results for males and females. This differentiation of the sexes may create a limiting effect on the career options that are available to one sex or the other and thus sex biasing may be introduced.

A tentative operational definition of sex bias has been adapted by the National Institute of Education, Career Education Program planning group as follows:

> "Within the context of career guidance,
> any factor that might influence a person
> to limit--or might cause others to limit
> --his or her consideration of a career solely
> the basis of gender."

The scope of this paper will be to review the major interest inventories and to explore the nuances and complexities of the technical aspects in the development of interest inventories, their item sampling, norming, scoring, reporting of results, and changing patterns of interests in relation to the differential treatment of sexes and to suggest guidelines to eliminate or alleviate any potential sex biasing factors.

Major Interest Inventories--a brief overview

Currently, there are two major interest inventories that are used to provide results that are helpful in vocational guidance:

1) the Kuder Occupational Interest Survey, and 2) the Strong Vocational Interest Blank. Both of these major inventories have handled the roles that males and females have in the occupational world in slightly different ways.

The Kuder Occupational Interest Survey (OIS) is intended for college-educated subjects (Kuder, 1966). It contains 100 items in a forced-choice format; for example, the subject is presented with a pattern of three possibilities, such as "Go to the movies", "Play cards", and "Go to a big party", and the subject is to pick one of three like most and one liked least or disliked. Results indicate the similarity of the person's interest in relation to satisfied subjects in a variety of occupations and college majors.

About 60 occupational-based scales and 30 college-major based scales are reported for the inventory. Male and female subjects respond to the same items. An earlier version of the Kuder, Form DD, the results for males were based on male criterion samples while scores reported for females were based on female criterion samples and for some selected male samples where there was not a female criterion sample. Thus, males and females were treated separately in scale development and reporting of the results.

Recently, the reporting of scores for the Kuder test has included all scores for all subjects (Kuder, 1974). Scores are reported for males based on male criterion groups and female criterion groups, and results for females are handled in an identical way. However, male and females are still treated separately during scale development.

The Strong Vocational Interest Blank (SVIB), like the Kuder, is geared more for professional occupations than non-professional occupations (Campbell, 1966; Campbell, 1971; Strong, 1943; Strong, 1959). Generally, each subject is asked to respond "Like", "Indifferent", or

"Dislike", to 400 items that cover a variety of areas such as occupational titles, activities, and amusements. Currently, about 55 occupational scales are reported, such as Mathematician, Chemist, Life Insurance Sales. In addition, there are about 20 basic interest scales, or homogeneous types of scales that measure interest preferences in broader terms, such as Mechanical, Teaching, Sales, and Sports. (The exact number of scales reported depends on the sex of the subject). The SVIB treats males and females distinctly from the very start with a separate set of test items for males and females. Although many items are identical between the make and female forms, about 40% of the items are unique to each sex.

Therefore, both the Kuder and Strong inventories may introduce a potential sex bias by a differential treatment of the sexes. The introduction of sex bias can occur at several different stages in development of an interest inventory: 1) in actual test items that appear on the inventory; 2) in construction of scales; and 3) in norming and reporting of results.

TEST ITEMS

The various forms of the Kuder inventories have one form to be used by both males and females, and thus all subjects are asked the same set of questions. There is a non-differential treatment of males and females at the item level.

The SVIB, however, is unique in offering a separate test booklet and different items for each sex. Although there is considerable overlap in the item content between the male and female versions of the SVIB, more than 1/3 of the test items are unique to each sex. For example, the male version of the SVIB, form T399, asks the subject to

respond "Like", "Indifferent", or "Dislike", to the following occupational possibilities: "Military Officer", "High School Principal", "Geologist", and "Public Relations Man", but the female version, TW398, omits these items. Likewise, females are asked to respond to "Supervisor in Telephone Office", "Stewardess", "Receptionist", and "Fashion Model" and these items are not in the male version.

There are both males and females gainfully employed in the above occupational examples--there are female military officers and male fashion models and these items should be acceptable to ask of either males or females. When different sets of items are asked of males and females, there may be the speculation that some types of activities and vocations are more appropriate for one sex and not the other. Doubt may be raised about the appropriateness of females "liking" to be in the military or having mechanical interests, or of males "liking" to do cooking, or teaching children, or being a flight attendant.

The SVIB has a long history of providing separate test forms of females and males. The first version of the SVIB was published in 1927 and was intended to help provide vocational guidance for males; six years later in 1933 the SVIB for females was published. During the 1930s there were differing employment patterns for men and women, most of the trades and professions had a heavy representation of males such as mechanics and dentists while women dominated the secretarial positions and elementary school teaching postions. Two forms were kept in existence for the last 40 years, primarily because of the differing patterns of males and females in some occupations and because there was demonstrable differences in response patterns of men and women to the same item.

The above factors sufficed to keep the SVIB male and female forms separated in the past but growing concern for equal rights and equal

opportunities and concern over possible career limitations has become of primary importance.

Recommendation

The recommended solution for avoiding any possibility of sex bias at the item level is to ask the same set of questions to both males and females. Special care should be taken so that the items are phrased so that they are not inherently referring to one gender; for example, Police Officer would be preferable to Policeman or Policewoman, Realtor would be better than Real Estate Salesman, Sales Clerk instead of Saleslady.

When gender can not be eliminate from the item, the item should include both possibilities; for example, Dressmaker/Tailor, Waitress/Waiter. Airline Flight Attendent would be better than Airline Steward/Stewardess from sexual bias standpoint, but perhaps not as easily understood by high school students (the reading level and comprehension of the items also must be a consideration if the inventory is to be applicable to high school students).

Test developers should give serious consideration to requesting the U.S.-GSA Civil Rights Office to recommend changes in item content so as to preclude bias toward any minority group.

While possible, with enough effort and care in writing items so that one gender is not inherently favored, this is perhaps the easiest solution in avoiding sex bias during the development of an interest inventory. Males and females do show different base rates of responding to the same item and this is elaborated below.

Male-Female Item Response Differences

Male and females do differ in their base rate of responding to interest items. For example, when responses to the SVIB item "Interior

Decorator" are investigated by gender, a substantial majority of women answer "Like" (67% of a sample of 1000 employed women representing an "average"). While, in contrast, a minority of men respond "Like" (28% of an "average" male sample, N=1000). Consequently, when a male responds "Like" to the item he is giving an unusual response and actually is responding in a manner similar to those of artists, actors, and architects. While a female responding "Like" to the same item is indicating little that is unique for her gender.

Similar types of male-female differences are found for the Kuder. For example, in responding to the item triad of "Go see a fire", "Go to see an accident in which people have been hurt", "Go see a famous person riding along the street", 57% of men-in-general indicate they would rather go see a fire in contrast to 33% of the women-in-general. While 60% of the women-in-general would prefer to see a famous person only 33% of the men-in-general would prefer that activity over the other two possibilities.

As indicated above, male-female differences do exist in their responses to the same item and as indicated below these differences are apparent during adolescence and are substantial.

Results from Teen-agers

Research on the differing patterning of interests between males and females by Campbell (1974) indicates that such differences appear even among early teen-agers. The following table lists occupational items that a majority of 8th grade males or females "Like" from a list of 130. With exception of occupations "Cartoonist" and "Professional Athlete", there is a definite difference in the responding of males and females even at this early age.

Insert Table 1 About Here

Table 1

Popular Occupations Among Male (N=81) and Female (N=76) Eighth Graders

Occupations marked "Like" by more than one-half of the boys

Occupation	Male Percentage "Like"	Female Percentage "Like"	Percentage Difference
Auto Racer	65%	24%	41%
Jet Pilot	57	22	35
Cartoonist	57	61	-4
Professional Athlete	53	45	8
Inventor	51	17	34

Occupations marked "Like" by more than one-half of the girls

Occupation	Female Percentage "Like"	Male Percentage "Like"	Percentage Difference
Children's Clothes Designer	76%	14%	62%
Interior Decorator	68	21	47
Fashion Model	66	11	55
Costume Designer	64	13	51
Steward/Stewardess	64	20	44
Actor/Actress	63	33	30
Home Economics Teacher	61	12	49
Cartoonist	61	57	4
Elementary Teacher	61	26	35
Nurse's Aide/Orderly	59	10	49

Table 1 continued

Manager, Child Care Center	58	11	47
Dressmaker/Tailor	57	15	42
Photographer	56	34	22
Manager, Women's Style Shop	55	10	45
Waiter/Waitress	55	16	39
Artist	54	35	19
Typist	53	9	44

As Table 1 indicates, there are considerable differences between young males and females. As they grow older, the magnitude of the differences does diminish somewhat, but never vanishes. The SVIB has had a Masculinity-Femininity (M-F) scale for many years to measure adult male versus adult female interests based on item responses, as have many personality tests such as the Minnesota Multiphasic Personality Inventory, (Welsh and Dahlstrom, 1956), and the California Psychological Inventory (Gough, 1969).

Results from Adults

With development of a female version for the SVIB in the 1930s Strong was able to measure the extent of different base rates of responding to interest items by males and females and developed M-F and F-M scales to measure those differences (Strong,1943). Research by Johansson in 1969 (reported in Campbell, 1971), further refined these scales.

Refined M-F and F-M scales were developed as part of a project studying male-female differences within occupation to the same item. When occupational membership was held constant, employed adult males ar females showed significant and practical differences. Fourteen occupations were available that had an adequate sampling (sample sizes about 250) of both males and females. All subjects indicated that they were satisfied with their jobs and all had been employed in their occupation for at least three years. Each occupation was represented equally. Thus, the samples were very precisely defined and potentially spurious results from comparable random samples of males and females was lessened.

This research was important in investigating the extent of similarity or dissimilarity between an average composite of males and females when samples from the same occupations were compared. Investigation

of these male-female differences within each specific occupation leads to three tentative hypotheses (Campbell, 1974):

1. Men and women in the same occupation do not differ in their interest preferences.
2. Men and women have different interests which are constant across all occupations.
3. Men and women have different interests which are specific to each occupation.

Table 2 below shows the number and percentage of items that showed large differences (15% or greater) between males and females within each occupation. Clearly there are considerable differences between males and females even when occupational membership is controlled; about 30% of the items in question showed large differences. Items that were investigated were the 229 items that were overlapping items between the male and female forms of the SVIB.

Insert Table 2 About Here

Those items that showed significant differences between the sexes were aggregated into the refined Masculinity-Femininity scale. Interests that were more typical of males than females included outdoor, adventuresome types of activities, and business interests. Interests more typical of females involved a liking for cultural activities, interest in fine arts, music, art, literature, and so forth.

To investigate whether differences are constant or unique to each occupation, the following table lists all items that showed very large differences (25% or greater) between the SVIB men-in-general (MIG; N=1000) and women-in-general samples (WIG; N=1000). For the first 14 items, females responded "Like" more frequently than males; for the

Table 2

Number of Items of 229 That Show "Like" or "Dislike" Response.

Differences of 15% or Greater Between Males and Females

Occupation	Male-Female Differences Number of Items	Percentage
Artists	60	26%
Bank Personnel	97	42%
Chemists	57	25%
English Teachers	73	32%
Interior Decorators	47	20%
Lawyers	67	29%
Life Insurance Sales	90	39%
Mathematicians	49	21%
Math-Science Teachers	93	41%
Medical Technologists	55	24%
News Reporters	76	33%
Physicians	61	27%
Psychologists	47	20%
Social Science Teachers	96	42%
Average	69	30%

last 5 items, male responded "Like" more frequently. Also, the differences between the percentage of males and females for the 14 occupations who responded "Like" to each item appear as the table entries. For example, 61% more of the females in the in-general sample prefer to "Decorate a room with flowers" than do males (75% of the women versus 14% of the men). The next entry indicates the difference between male and female artists, next for bank personnel, and so forth. The last entry indicates the average across the 14 occupational samples.(Campbell, 1974).

Insert Table 3 About Here

Inspection of the first and last columns indicates the entries are very similar, differing by only one or two percentage points. Thus, as the first column indicated, men and women in a "general" sample do have different response patterns to the same item and the last column indicates that these differences are fairly constant between males and females even when occupational membership is held constant.

The above illustrated items that showed very large differences between the sexes; approximately one-half of the 325 items on the revised SVIB (that will be released in 1974) indicate differences of 15% or more between males and females comprising the in-general samples. (The increase over the 30% for the 229 in Table 2 is a result of including items relating to domestic, mechanical, and military types of activities that previously appeared on only one of the SVIB forms and these types of activities show large differences).

Thus, even though good items can be written for males and females to alleviate potential sex bias by not inherently referring to one gender; different response patterns by sex are a concern for scale

Table 3

SVIB Items Showing Very Large Differences Between Men and Women in the same Occupations

(Table entries are the percentage difference in "Like" response to the designated item for the two sexes.)

Items Favored by Women	Bus. vs. Mfg	Artists	Bank Personnel	Chemists	English Teachers	Interior Decorators	Lawyers	Life Insurance Sales	Mathematicians	Math-Science Teachers	Medical Technologists	News Reporters	Psychologists	Social Science Teachers	Physicians	AVERAGE
Decorate a room with flowers	61	49	64	55	57	23	54	62	43	59	55	66	59	62	58	55
Interior Decorator	47	37	43	35	43	2	43	46	23	35	41	54	45	48	40	41
Work with babies	28	15	34	19	19	33	12	35	36	19	12	19	18	18	13	22
Religious music	27	20	38	27	19	16	29	19	12	29	14	25	20	27	28	23
Magazines about art & music	27	9	31	22	23	11	11	22	16	21	21	29	28	37	26	24
Private Secretary	26	3	51	11	27	0	20	37	16	34	12	18	22	28	16	22
Plan a large party	26	22	22	22	24	17	24	22	19	23	7	33	18	29	11	21
Work with ballet dancers	25	17	29	13	25	18	23	38	19	19	23	28	25	28	25	24
Interpreter	25	10	25	25	22	13	29	19	18	18	15	27	24	29	15	21
Items Favored by Men																
Travel alone	33	31	42	23	46	29	40	31	30	49	35	40	27	41	42	36
Popular mechanics magazines	33	25	25	18	36	20	29	28	11	24	29	30	24	40	25	2[illegible]

Table 7 continued

Repair electrical wiring	29	20	33	17	39	21	24	33	17	34	28	24	27	31	25	27
State governor	25	21	24	13	30	10	13	19	27	32	32	30	12	25	25	23
Airplane pilot	25	32	34	12	22	13	19	30	14	35	13	18	19	25	19	22

construction and reporting of results.

INTEREST INVENTORY SCALES

Another stage at which sex bias may become apparent is during the construction and norming of scales for an interest inventory. Two main types of scales frequently are developed: 1) criterion scales where a person's interest preferences are associated with those in a criterion sample such as an occupational sample of artists or college students majoring in biological sciences; and 2) homogeneous or basic interest scales that measure the underlying interest dimensions of the inventory and a person's interest preferences are related to those in a general reference sample. Each has its unique advantages in the type of information that it captures from items on the inventory and each differs in the way that potential sex bias may be introduced.

Criterion Scales

Strong

The development of criterion or occupational scales for the SVIB involves the contrasting of item responses of employed people in a specific occupation (criterion sample) with a sample representing a population "average" (in-general sample). Each occupational scale contains those items that significantly differentiate criterion sample from in-general sample. There are various methodological considerations and problems that have to be considered in collecting a criterion sample and defining an in-general sample (see Campbell, 1971; Clark and Campbell, 1965; Strong, 1943), but for purposes of this paper the

assumption is that adequate samples are available.

Therefore, after item responses for the criterion sample and in-general sample have been collected, item response percentages for both groups are calculated and contrasted with each other. Items that show significant percentage differences (usually 15%-20%) between the two groups are included in the occupational scale. Response patterns (for example, Like, Indifferent, or Dislike) for these significant items typically are assigned weights based on the magnitude and direction of differences between criterion sample and in-general sample. The resultant scale is then normed on the criterion sample used for scale development.

To illustrate, the following example shows the response percentages of male psychologists and the male in-general sample for the two items "Author of Novel" and "Employment Manager."

"Author of Novel"

Response	Male Psychologists (N=252)	Men-in-General (N=1000)	Response Differences	Response Weights
Like	81%	51%	30%	+1
Indifferent	15%	30%	-15%	-1
Dislike	4%	19%	-15%	-1

"Employment Manager"

Like	35%	29%	6%	0
Indifferent	34%	41%	-7%	0
Dislike	31%	30%	1%	0

As illustrated in the above examples, responses are empirically unit weighted if response difference; indicate that there is a significantly large difference between criterion sample and in-general sample; responses are not weighted if the difference is small. [After years

of research with the SVIB, a 15% difference or greater is considered large and meaningful; Campbell, 1971]. Thus, if a person responds "Like" to the item "Author of a Novel" his or her raw score on the psychologist scale would be incremented by one, if the response was "Indifferent" or "Dislike", the raw score would be decremented by one. Responses to the item "Employment Manager" would have no bearing on the score for the psychologist scale, but it would be relevant for some other scales such as the sales mamager scale.

The number of items that differentiate an occupational sample from the in-general sample on the SVIB is usually between 70-90 items or about one-fourth of the test. A person's raw score on each scale developed for the inventory is the sum of plus and minus unit weights that correspond to his or her unique pattern of responses for each scale. The raw score is then converted to a standard scord using a raw-score-to standard-score conversion formula. Each person receives a standard score for each of the occupational scales developed for the inventory indicating the degree of similarity betwee his or her interest preferences and those in the criterion sample based on those items that were more characteristic of the criterion sample than the in-general sample.

Kuder

The Kuder OIS uses a somewhat different approach than the SVIB in the development of empirical criterion scales for the inventory. Because of difficulties found by Campbell and Strong in developing an adequate in-general sample, Kuder (1966) decided to circumvent the methodological problems by use of a lambda coefficient which is similar in concept to biserial correlation coefficient. Lambda coefficient expresses the degree of similarity (correlation) between a subject's

responses and those members of the criterion group, such as an occupational sample. This procedure effectively eliminates the need for an in-general sample. The upper limit of lambda is 1.00 indicating complete similarity with responses of those in the criterion sample, a lambda of .00 indicates no similarity; these coefficients are used as the scores for the Kuder Occupational Interest Survey. Because of the method of scale construction, norming and standard score transformations are not a concern as with the SVIB.

Gender as a Factor

Since the SVIB traditionally has separated the response patterns of males and females at the item level, so too are the responses separated at the scale development level. Male occupational samples are compared with a male in-general sample representing the "average" employed male, and female occupational samples are compared with a female in-general sample representing the "average" employed female. Thus, it is not surprising to find that if a person completes both the male and female SVIB, different scores will be obtained for the same-named scale on both inventories; for example, a person may obtain a score of 45 on the male chemist scale and 40 on the female chemist scale. Item content of the like-named scales are different, criterion samples are composed of different people, and the in-general samples are different.

The Kuder also separates the sexes during scale development. Some occupations such as Lawyers and Computer Programmers have separate and distinct scales for males and females. Other criterion scales such as Dean of Women or Plumber are based on just one sex. Currently, a subject is scored on all scales regardless of sex and the same problem arises as with the two forms of the SVIB--where two scales exist for an occupation,

one based on females and one on males, discrepant scores result.

The sex bias concern is that there is not complete comparability in the scales that are developed for male and female forms of the SVIB and scales developed by gender for the Kuder. The female SVIB contains scales that measure interest preferences for various nursing occupations, airline stewardesses, entertainers, and telephone operators. However, the male SVIB does not give an indication of these specific occupational interests--there are adequate samples available for male airline stewards, male nurses, male entertainers, and male telephone operators (Schlossberg and Goodman, 1972). The male SVIB has a wider sampling of professional occupations such as biologists, architects, psychiatrists, and physicists but these are not available on the female SVIB. A similar set of circumstances exists for the Kuder.

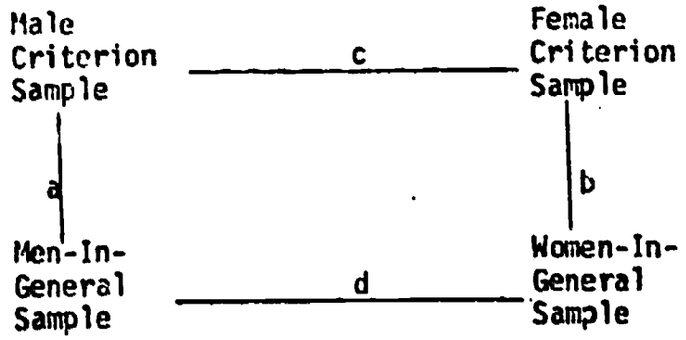
For someone concerned about sex bias in interest measurement, separation by gender when scales are developed and reported is a salient focal point. For example, are interests measured by the male form of the SVIB different in scope than those measured by the female form? How important are these differences? If a woman wants the results of scales appearing on just the male form then there is the problem of how generalizable are the results from male criterion groups to her interest preferences? This is also true of the Kuder, but it is perhaps now more apparent.

Subjects completing the Kuder are now presented with two scores for some occupations; with the SVIB, the subject had to take two tests to produce differing results. The following section investigates occupational scale characteristics when they are constructed on samples separated by gender.

Sexual Stereotypes in Criterion Scales

The relative importance of sex differentiating items in like-named scales developed on male criterion samples and female criterion samples is contained in research by Johansson and Harmon (1972) on the SVIB. The study specifically investigated the following areas: 1)whether men and women in the same occupation have different interests; 2) whether these differences are reflected in occupational scales for men and women; and 3) whether these occupational-scale differences are valid and useful.

Fourteen occupational samples and the in-general samples were used for analysis as discussed previously (see pages 10-11). The following figure presents a paradigm of how item-response differences were analyzed using male and female criterion samples and their relationships to the in-general samples.



The male occupational scale is based only on differences between the male criterion sample and men-in-general (Difference a). The female scale is based only on differences between the female criterion sample and women-in-general (Difference b). Difference c represents items that differentiate male-female criterion samples and examples of this have been given previously in Table 1. For those items that show response differences between men and women in an occupation (Difference c),

a given item can have four possible effects on male and female scales.

1. Male-female differences are not incorporated in the scale.

Despite a large item-response difference (for purposes of this study, large was defined as 12% or greater) between males and females (Difference c), the item appears on both scales because differences between occupational and in-general groups (Differences a and b) are also large. Thus, sex difference is not differentially incorporated into the scales. The following exemplifies this condition where there are large response differences to the item "Electronics Technician" between male and female Medical Technologists, between in-general samples, and between criterion samples and same sex in-general samples. Even though the item is a sexually stereotypic item (a large difference between males and females), it is weighted the same for both male and female scales, and does not produce a sex bias in the scale content.

Non-Influential Male-Female Differences

Reponse Percentages to the Item:

"Electronics Technician"

	Medical Technologists			In-General Samples			Response Weights	
Response	Male	Female	Differences	Men	Women	Differences	Male Scale	Female Scale
Like	55%	35%	20%	22%	11%	11%	+1	+1
Indifferent	35%	36%	1%	42%	30%	12%	0	0
Dislike	10%	29%	-19%	36%	59%	-23%	-1	-1

Across all fourteen occupational samples, approximately one-fifth of the items differentiated males from females in the criterion samples and were non-influential because they were weighted identically on the male and female scales.

2. Valid male-female differences are incorporated in the scales.

While there is a large difference between the criterion groups (Difference c), only one of the differences between the criterion groups and the same sex in-general sample (Difference a or b) is also large; thus the item appears on only one of the occupational scales. In this situation, Difference d in the paradigm is small. Example 2 below shows that responses to the item "Computer Operator" are weighted on the male scale for Medical Technologists but not for the female scale. Differences between male and female Medical Technologists are large but this difference is specific to the occupation since the in-general samples show a much smaller difference.

Valid Male-Female Differences

Response Percentages to the Item:

"Computer Operator"

	Medical Technologists			In-General Samples			Response Weights	
Response	Male	Female	Differences	Men	Women	Differences	Male Scale	Female Scale
Like	31%	28%	3%	14%	22%	-8%	+1	0
Indifferent	51%	34%	17%	37%	29%	8%	+1	0
Dislike	18%	38%	-20%	49%	49%	0%	-1	0

The above item illustrates the inclusion in scales of a valid or legitimate sex difference in interests between males and females in the same occupation. Averaged across all 14 occupational samples, approximately 10% of the scale items fell into this category.

3. Non-valid male-female differentiating items incorporated in the scales.

Items that differentiated the two in-general samples (Difference d) were regarded as sexually stereotypic since it was true of all males and females and for males and females in the criterion samples. When such a sexually stereotypic item appeared on only one scale, then it seemed to be spuriously related to the magnitude on the difference between criterion sample and in-general sample. The example below shows this relationship for the item "Art Galleries." Males and females in the criterion samples and in-general samples show large differences but the item is only weighted on the male psychologist scale.

Non-valid Male-Female Differentiating Item

Response Percentages to the Item:

"Art Galleries"

	Psychologists			In-General Samples			Response Weights	
Response	Male	Female	Differences	Men	Women	Differences	Male Scale	Female Scale
Like	66%	86%	-20%	42%	66%	-24%	+1	0
Indifferent	25%	11%	14%	39%	24%	15%	-1	0
Dislike	9%	3%	6%	19%	10%	9%	-1	0

While items as exemplified above did represent significant differences between males and females in the criterion samples and the in-general samples, the items were related to sexual stereotypes that had been included fortuitously in the scale for one of the sexes. Fortunately, only a few items of this type are included in the scales.

4. Non-valid in-general differentiating items incorporated in the scales.

Again, differences between the in-general samples (Difference d) were considered sexually stereotypic. If the item showed a small difference between the criterion samples (Difference c), but was included in one of the scales because a large difference between a

criterion sample and in-general sample (Difference a or b), another type of invalid difference had been included in the scales. As illustrated below, male and female psychologists respond similarily to the item "Inventor" but the in-general samples show large differences. The item is weighted on the female psychologists scale but not on the male scale. The male and female scales differ, not because people in the occupation differ, but because in-general samples differ.

Non-Valid In-General Differentiating Items

Response Percentages to the Item:

"Inventor"

	Psychologists			In-General Samples			Response Weights	
Response	Male	Female	Differences	Men	Women	Differences	Male Scale	Female Scale
Like	73%	71%	2%	62%	44%	18%	0	+1
Indifferent	22%	22%	0%	28%	35%	-7%	0	0
Dislike	5%	7%	-2%	10%	21%	-11%	0	-1

These last two types of differences (#3 and #4), influencing one scale and not the other, did not represent a valid inclusion of the item in the scale for one sex and not for the other. Across all 14 occupational scales that were studied, about 18% of the items on the scales were these types of items that represented the inclusion of non-valid sex differences in the scales.

In summary, the study showed more than 70% of the items on the SVIB scales studied did not differentially incorporate sexually stereotypic items into male and female scales. Of the remaining items on the scale, about 10% were items reflecting valid sex differences and less than 20% were invalid inclusions. Thus, depending on one's point of view, 70% can be taken as the over-riding factor with little worry about sex bias or 20% can be glared at as justification for

change.

Solutions to the Problem

As the preceding data have indicated, there were differential base rates of responding for the sexes to individual items on an interest inventory. As will be shown later, these differences have remained stable through the late 1960s and probably will be apparent for years to come. These differences can not be ignored and become a vexing problem when occupational scales are developed. There are four possible strategies one can use in handling this problen as outlined by Campbell (1974):

1. A simple solution would be to ignore sex and randomly collect criterion samples and let the proportions of males and females in the sample match the population split. There would be two attendant disadvantages to this approach. One, some occupations still have a low percentage of one sex in the occupation, for example, female carpenters, and collecting a truly random sample with sufficient representation of both sexes would be impossible. Two, if occupational scales are constructed by contrasting a criterion group with an "average" sample, such as an in-general sample, the composition of the in-general sample would over-represent those occupations where there is a preponderance of one of the sexes.
2. A second possibility would be to use an equal representation of males and females in the composition of criterion groups, again developing just one scale for both sexes. As in the previous alternative, adequate samples of both sexes are extremely difficult to collect and almost impossible for some occupations. Where there are sufficient members of an occupation, male response preferences could be statistically weighted equally with female

response preferences. However, for some occupations, a low representation of one sex would lead to instability in the combined responses for the total group. This procedure would wash-out items that are valid for one sex but not for the other sex. for example, "Decorating a room with flowers" is more important if a male responds "Like" than if a female responds "Like".

3. A third possibilty would be to develop separate scales for males and females and then to equate scores through a statistical formula appropriate for each occupation. This would be analogous to procedures used be some graduate schools in applying differential weights to grade-point averages obtained at various undergraduate institutions, with "heavier weights" corresponding to the more academically "hard" institutions. However, considerable research would have to be done on such modifications of scores to investigate the impact of such a procedure. The end result probably would be the same as the second alternative above.
4. A fourth possibility would be to develop separate scales for males and females as is currently done for the Kuder and SVIB. This is the easiest to do from a developmental standpoint and yields the best predictive and concurrent validity for the inventory. Studies by Berdie (1961), Dunnette and Kirchner (1958), Frederiksen, Melville, and Gilbert (1954,1960), Ghiselli (1963), Grooms and Endler (1960), Johnson and Johansson (1972), and Seashore (1961), have shown the efficacy of moderator variables in increasing the validity of measurement. Using sex as a moderator variable increases the validity of scales by including those items that are most differentiating for each sex.

Recommendation

Collection of occupational samples is a costly venture, but the

demonstrated predictive validity of empirical scales developed on criterion samples makes the utility of this approach worthwile. Wherever feasible, a concerted effort should be made by the test developer to collect adequate samples of each sex for an occupation. Construction of separate scales for males and females is technically sound from the standpoint of concurrent validity. However, it is recommended that greater care be given so that non-valid sex differentiating items (as outlined in pages 24-26) are not incorporated into the scales.

Use of Criterion Scales Developed on the Opposite Sex

Currently, scores are reported for subjects based on opposite-sex criterion samples for the Kuder and for the SVIB if the opposite-sex test is administered. This procedure will continue until adequate representation of both males and females in all occupations is accomplished. The utility of this procedure has been explored in various research studies.

Research by Cole (1973) on female interests and Cole and Hanson (1971) on male interests suggests that there are similarities between occupational configurations of women's interests and configurations of men's interests using data on the SVIB, Kuder, and two other inventories. These configurations could be useful in providing additional information about career opportunities for males and females even though there are not relevant specific scales on the inventory. The aboye research indicates that there is enough similarity in interest structure between the sexes that generalizations beyond the status quo of an inventory are possible in exploring new career opportunities for males and females. While these findings are important in understanding configuration of interests, test users will find occupational scores the easiest to use and there still is the problem of reporting scores for females based on male criterion groups and reporting

scores for males based on female criterion groups.

Applicability of occupational scales developed on male criterion groups for females taking the Kuder OIS has been studied by Hornodary and Kuder (1961). They found that scales differentiated for women as well as they did for males for nine of ten scales studied. In a similar study Kuder (1966) found high median correlations between scores based on male criterion samples and female criterion samples for three samples of women. Kuder concluded that reporting of scores for females based on male criterion samples was a valid procedure for representing their in fields where there are opportunities for women but criterion data was not available. Kuder (1966) also stated that if a woman enters an occupation dominated by males, she will find greater satisfaction if her interests resemble those of the males in the occupation.

More recently, Kuder (1974) stresses the importance of giving more emphasis to scores based on the subject's own sex and uses scores based on the opposite-sex criterion scales to give added insight into a subject's interests. If the person scores highest on several of opposite-sex criterion scales, these scores may indicate good possibilities for further exploration if not represented by same-sex scales.

Darley and Hagenah (1955) found that using both forms of the SVIB to be beneficial for women who had a high degree of maturity, ability, and career motivation. Strong (1959) suggested that the male form could be used for women who have interests (via the M-F scale) and career aspirations similar to those of males. Thus there is some evidence on the efficacy of reporting scores for scales based on criterion samples that are of the opposite sex of test taker.

A study by Laime and Zytowski (1964) also indicated that for women who took both forms of the SVIB, their score on several males scales could be predicted from corresponding female scales. In

addition, for more professionally oriented females, they tended to receive higher scores on certain scales of the male form than on the female form.

Huth (1973) cites criticisms of the SVIB as it pertains to measurement of female interests. She generally concluded as did Super and Crites (1962) ten years earlier, that the female SVIB does not show good differentiation of interests for the majority of women. There seemed to be a commonous of interests among women that makes differentiation difficult except in those cases where women have a clear-cut interest. The general pattern seems to be a "home versus career" orientation. Thus, if women have had a strong career orientation, the male SVIB has been used frequently to provide the necessary differentiation of occupational interests.

A more extensive analysis of the same type of data was done by Campbell (1974). Using only items that were common to male and female SVIB, 31 male and 31 female occupational scales were developed and normed on appropriate-sex norm groups. For example, a male Physical Therapist scale was developed and normed on male physical therapists; likewise, a female Physical Therapist scale was developed and normed on female physical therapists. Next, a general sample of 200 males and 201 females were scored on the 62 scales (31 male-based scales and 31 female -based scales), correlations were computed between the two scales for each occupation and mean differences in score value were derived. Table 4 presents the results of this study.

Insert Table 4 About Here

The median correlation between same-named scales for the SVIB reported in Table 4 ($r=.76$ and .77) was equal to the magnitude obtained

Table 4

Results of Using Male and Female Scales with Male and Female Samples

	Male Sample				Female Sample			
Scale	Correlation between M & F Scale	Male Scale Mean	Female Scale Mean	Mean Difference	Correlation between M & F Scale	Female Scale Mean	Male Scale Mean	Mean Difference
Physical Therapist	.87	33	37	- 4	.90	35	27	+ 8
Army Officer	.65	33	45	-12	.63	31	21	+10
Army Non-Commissioned Officer	.73	33	43	-10	.71	31	21	+10
Engineer	.90	38	38	0	.89	20	25	- 5
Mathematician	.79	24	30	- 6	.77	17	26	- 9
Computer Programmer	.91	37	38	- 1	.88	24	26	- 2
Math-Science Teacher	.64	34	39	- 5	.64	30	27	+ 3
Medical Technologist	.85	26	36	-10	.89	26	20	+ 6
Dentist	.61	31	33	- 2	.76	24	31	- 7
Physician	.76	29	37	- 8	.81	25	24	+ 1
Psychologist	.82	29	31	- 2	.35	21	34	-13
College Professor	.73	36	42	- 6	.75	33	41	- 8
Artist	.81	25	28	- 3	.83	26	35	- 9
Interior Decorator	.72	24	15	+ 9	.68	17	39	-22

Table 4 continued

Musician	.56	34	25	+ 9	.59	36	46	-10
English Teacher	.39	30	23	+ 7	.90	33	40	- 7
Librarian	.75	23	30	- 7	.73	31	38	- 7
Reporter	.64	28	32	- 4	.85	31	38	- 7
Social Worker	.91	25	27	- 2	.90	29	31	- 2
Recreation Leader	.94	29	33	- 4	.92	35	27	+ 8
Elementary Teacher	.57	34	23	+11	.46	35	39	- 4
Music Teacher	.75	23	18	+ 5	.71	27	39	-12
Guidance Counselor	.83	30	27	+ 3	.78	31	31	0
Social Science Teacher	.74	33	34	- 1	.72	33	34	- 1
Lawyer	.71	31	38	- 7	.73	28	33	- 5
Life Insurance Sales	.83	23	28	- 5	.75	24	20	+ 4
Buyer	.76	26	27	- 1	.75	24	24	0
Pharmacist	.46	29	34	- 5	.58	27	26	+ 1
Accountant	.77	23	32	- 9	.78	20	9	+11
Banker	.76	30	33	- 3	.73	29	21	+ 8
Business Education Teacher	.74	31	20	+11	.69	21	29	- 8
AVERAGES	.75			5.5	.77			6.7

by similar studies on the Kuder OIS (Kuder, 1966). The magnitude of the correlations would indicate that there is a great deal of similarity between scales developed on male criterion groups and female criterion groups, but not sufficient similarity to interchange scales carte blanche.

Males in the general sample were scored on both sets of scales for each occupation as were the females. The mean difference column entries are important in investigating the scores for females on scales developed on female criterion samples. Of importance is the direction of the difference between the male scale means and female scale means. Since the means are for random samples of males and females, low scores on occupational scales are desireable; negative differences indicate that appropriate-sex scales are superior and positive differences indicate the obverse. The results indicate that scales based on male criterion groups work best for males, and scales based on female criterion groups work best for females.

The average absolute difference between the two scales for each occupation in Table 4 was about one-half standard deviation(5.5 standard score units for males and 6.7 for females). If there were no overall differences in scales developed on male and female criterion samples, then the difference would be about zero. This clearly was not the case, mean differences of the magnitude of one-half standard deviation on SVIB occupational scales are both statistical and meaningful differences.

In investigating item composition of the above scales, Campbell found that scales that contained sexual stereotypic items influenced the magnitude of the score differential between males and females. Males tended to score highest on female scales dominated by "male" types of items, such as the female scales for Army Non-Commissioned

Officer and Army Officer, than on the corresponding male scale, while females scored highest on male scales dominated by "female" types of items, such as the male scale for College Professor, Musician, and English Teacher.

Recommendation

So as not to limit career opportunities available to males and females, it is recommended that scores be reported for all scales available and when possible based on the appropriate sex. When scores based on the opposite sex are reported, it is imperative that test users be made cogently aware of this. When these opposite-sex scores are used, they should be interpreted with sexual stereotypes, and their potential effect on scores, kept in mind.

INTERNALLY BASED SCALES

The second major type of interest scale developed for interest inventories is based on internally related items and labeled as homogeneous scales or basic interest scales. Generally, each inventory, through its item content, covers a range of basic interest dimensions and scales can be constructed through statistical procedures to measure these dimensions, for example, Mechanical interests, Social Service interests, and Numerical interests. Unlike the occupational scales where the number of scales developed depends on occupations that test developers are willing to spend time and money to test, the homogenous scales are internally-based and developed to tap all interest domains of the inventory. The male SVIB has 22 homogenous scales and the female SVIB has 19. The Kuder OIS does not contain these types of scales, but a version intended for junior high school students, Kuder General Interest Survey, has 10 such scales.

Scale Development

Typically, a general sample is tested with the inventory and product-moment correlation coefficients are computed for all pairwise combinations of items by assigning numeric weights to response patterns. Then using factor analsis or cluster analysis, highly interrelated items are aggregated into a scale and a descriptive name applied that reflects some common psychological theme that is being measured by the individual items.

For example, the item responses of "Like", "Indifferent", and "Dislike", on the SVIB were assigned weights of +1,0,-1 respectively and then all pair-wise item intercorrelations were computed using the responses of a general sample for the male form, and a general sample for the female form. The example below shows a correlation of r=.63 between the items "Algebra" and "Arithmetic" based on the responses of 500 males. (The table entries are response percentages of a general sample of employed males to two items considered as a pair; for example, 51% of the subjects responded "Like" to both "Algebra" and "Arithmetic").

		"Algebra"				
		Like	Indifferent	Dislike	Total	
	Like	51%	12%	6%	69%	
"Arithmetic"	Indifferent	3%	11%	6%	19%	r=.63
	Dislike	1%	1%	10%	12%	
	Total	55%	24%	22%	100%	

These two items, "Algebra" and "Arithmetic", in addition to the items "Calculus", "Geometry", "Mathematics", and "Physics" appear on the basic interest scale MATHEMATICS for the male SVIB. The same procedure was used for the SVIB female basic interest scales using a general sample of female responses. As with the occupational scales

the person's raw score on each scale is the sum of the response weights for the items included on the scale.

Again, because the SVIB has used separate sets of items for males and females, separate basic interest scales have resulted with many overlapping concepts, but still some unique scales appear. For example, males are not scored on a DOMESTIC/HOMEMAKING scale nor are females scored on a MILITARY ACTIVITIES scale. (This uniqueness is a direct result of not having sufficient appropriate item contents on the inventory.) The previous recommendation that the same set of items be given to both sexes would eliminate this uniqueness in generation of results.

The ten homogenous scales for the secondary school version of the Kuder has identical homogenous scales for males and females and permits meaningful comparison between the sexes on their interest preferences. Using the same set of items for all subjects, the Kuder has overcome some of the objections of the SVIB.

Use of intercorrelations as the basis for scale development circumvents many problems associated with development of occupational scales. If the same set of items is administered to both males and females, it is recommended that a general sample with equal representation of both sexes be used in the construction of these types of scales. The predictive validity of internally-based scales is of a relatively lesser importance than for criterion scales and differential rates of responding for males and females are not a real problem until scales are normed.

Norming of Internally-Based Scales

The reference sample used for norming homogeneous scales is determined by the test developer and generally is influenced by which reference sample would be most relevant for interpreting the results.

If a general reference sample is used where males and females are equally represented, there are two alternatives to norming of scale and reporting of results.

One alternative would be to split the general reference sample or other reference sample by gender and provide separate norms for each sex so that the average male sample and average female sample would have identical means and standard deviations. Such statistical manipulation to achieve equality of score distributions would actually disguise the underlying male-female differences. For example, a raw score of 10 on the ART scale may convert to a standard score of 50 based on a male reference sample but the same raw score of 10 for a female may convert to a standard score of 45. The end result is that the same degree of interest would produce different standard scores. (The separation of the sexes when constructing empirical occupational scales can be faulted by the same argument. However, the intercorrelational procedure permits the combining of male and female responses without "washing out" differential response rates and the predictive validity of homogeneous scales in not an overriding concern. Both these factors make the use of a combined male-female sample more acceptable for internally-based scales.)

The second alternative is to develop one scale for both males and females using the general reference sample, norm the scale on the male-female composite, and provide normative data separated by gender when results are reported. This procedure would convert a raw score to the same standard scores regardless of gender, but interpretive norms would provide necessary information about differences in male-female score distributions. This procedure, illustrated below is similar to that for reporting scores on the Comparative Guidance Placement Ptogram of the College Entrance Examination Board.

Recommendation

The following example illustrates reporting of scores on two internally-based types of scales, Mathmatics and Physical Science using male-female composite for norming purposes but the interpretive data are split by gender. Standard scores are plotted and easy reference to both male and female distributions is possible. Explanatory statements can be preprinted on the profile explaining that the white bar represents the normal range of male scores and the speckled bar reprsents the normal range of female scores with vertical lines representing the average. Good graphic representation could also indicate quartiles or 10th and 90th percentiles.

Example of Homogeneous Scale Output

Scale	Score	0.....20.....40.....60
Mathematics	55	X
Physical Science	51	X

This procedure also would have the attendant advantage of providing a more meaningful linkage to scores of male and female occupational samples. Since a standard score would represent an equivalent degree of interest to items on the scale regardless of gender, resultant score presentations for occupations (identified by sex) would give meaningful comparisons between males and females in the same occupation.

The above recommended procedure uses the same items for both males and females, yields identical standard scores from the same expressed interests on homogeneous scales regardless of gender, but takes into account differential response frequencies for each sex when scores are related to normative data.

Sex Differences

Table 5 shows mean raw score differences between males and females on the homogeneous-type scores for the Kuder General Interest Survey (Kuder,1973). All differences between the sexes were statistically significant. Males scored much higher on the Mechanical scale (more than 19 points) than females, while females had more intense interests than males in Social Service. Generally, the results pattern sexually stereotypic roles, males scoring higher on Outdoor, Mechanical, Computational, Scientific, and Persuasive, while females were higher than males on Artistic, Literary, Musical, Social Service, and Clerical.

Insert Table 5 About Here

CHANGE OF INTERESTS OVER TIME

Do interest preferences change over time? For individuals, there will be a wide diversity in amount of change and this will depend on his or her age when tested. As the person becomes older, more and more experiences are encountered which will predispose the person to like or dislike that activity more than previously depending on the experienced reward value. Roughly one-third of the person's change in interest will occur prior to graduating from high school, one-third during college years (18-22), and one-third during adulthood with a great deal of stability after the age of 25 (Johansson and Campbell,1971).

Although considerable evidence exists indicating that people will change in their vocational interest preferences, do occupational group characteristics change sufficiently to invalidate tests constructed

Table 5

Means and Standard Deviations for Males and Females on Homogeneous Scales of Kuder

	Students in Grades 6-8				Students in Grades 9-12			
	Males (N=287)		Females (N=433)		Males (N=471)		Females (N=691)	
Scales	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Outdoor	27.45	8.73	22.93	8.20	24.50	9.71	21.56	8.39
Mechanical	43.13	10.20	26.01	9.24	44.74	11.67	26.31	8.84
Computational	30.03	8.34	27.62	7.80	32.99	9.84	27.46	9.79
Scientific	39.33	11.54	29.94	10.87	39.01	11.71	29.26	10.90
Persuasive	54.90	12.01	49.53	11.96	55.60	13.15	51.57	12.54
Artistic	29.73	8.82	34.47	9.49	29.13	9.35	33.71	9.86
Literary	29.41	9.44	35.71	9.34	30.34	10.23	34.49	10.68
Musical	10.67	7.51	12.33	8.17	11.75	8.20	13.58	7.50
Social Service	46.85	12.45	59.35	12.64	44.34	13.01	60.40	13.84
Clerical	53.52	10.13	61.76	12.45	52.10	12.91	61.04	15.47

5, 10, 20, 30 years ago? Specifically 1) are there general changes across various groups of people over time? 2) are these changes associated with the measurement of vocational preferences for specific occupational groups?, and 3) are these changes on items that currently differentiate the sexes?

Changes in Item Popularity

To investigate if there were general changes in interests, occupational samples tested in the 1930s were contrasted with a matching sample from the same occupation tested in the 1960s (Campbell, 1971). Items that showed a significant change in the base rate of responding across a majority of the occupational samples were identified on both male and female SVIB. For example, about 10% of bankers tested in 1934 responded "Like" to the item "College Professor", while about 30% of bankers tested in 1964 responded "Like". Approximately one-fifth of the items investigated showed large shifts in popularity over the 30 year time span. These types of items were used to develop a Cultural Change Scale; the scale correlated in the .80s with test year of the sample. Thus a strong relationship existed between year tested and items comprising the Cultural Change scale. Investigation of items indicated that recently tested samples had a more positive liking for activities that were extroverted in content, and concurrently out-doors and skilled trades activities became less popular. Results were similar for both males and females.

Changes in Group Characteristics

Since items could be identified that showed significant changes in popularity over time, the next question was whether or not these changes also were incorporated into occupational and homogeneous scales that measured the interest preferences of criterion groups. If occupational interest patterns do change with time, then interest

inventories should be revised continually. To investigate this problem, Campbell (1966) went back to some of the SVIB's original criterion groups and tested the men in the 1960s who were holding the exact same jobs as those tested by E.K. Strong in the 1930s. Four samplings were done: 1) 1965 ministers were matched with ministers in the same church in 1927; 2) 1964 bankers were matched with men in the same bank position in 1934; 3) corporation presidents tested in 1965 were matched with presidents of the same company in 1935; and 4) 1965 school superintendents were matched with superintendents of the same school system in 1930. The following table presents the data of 12 experimental homogeneous scales (Johansson, 1969,1974) that permitted comparison between the 1930 and 1960 testings (the SVIB was revised in the middle 1960s, precluding an exact comparison on all but 12 homogeneous scales). As is apparent in Table 6, the means between the two testings were very similar--there were greater differences among the different occupations than between two samplings of the same occupation.

Insert Table 6 About Here

Inspection of SVIB criterion scores for two samplings of each of four occupational groups revealed the same results (Campbell,1966). The striking similarity of results for each of the comparisons suggested that occupational scales developed in the 1930s were still relevant for the 1960s. Other research by Thrush and King (1965) on medical students, by Ferguson (1958, 1960) on life insurance salesmen, by Campbell (1965,1968) on male and female psychologists, and by Mayarazzo, et. al. (1964) on policemen and firemen clearly indicated that occupational preference data do not show dramatic shifts over time.

Table 6

Means for Four Occupations

Where the Job Was Held Constant Over Two Testings[1]

	Ministers (N=98)		Bankers (N=93)		School Supts. (N=149)		Corp. Pres. (N=25)	
Scale	1927	1965	1934	1964	1930	1965	1935	1965
Office Practices	50	49	61	61	51	51	43	45
Merchandising	46	46	49	48	44	50	45	45
Bus. Management	52	54	54	56	54	59	50	52
Law/Politics	49	51	51	49	52	54	47	51
Teaching	55	58	39	44	57	59	42	49
Art	49	49	41	41	43	43	42	41
Writing	50	53	40	41	46	48	40	42
Sports	49	47	46	48	44	49	42	43
Mechanical	53	50	51	53	50	51	55	53
Physical Science	53	49	46	46	53	52	53	50
Numerical	51	49	52	54	54	55	57	57
Medical Service	50	49	47	47	46	50	43	45

[1]Note--standard deviations vary around 9.

However, there was a slight trend in the Campbell(1966) study for the more recently tested occupations to score higher on more recently developed occupational scales which was attributable to using an in-general sample that spanned testings over some 40 years and this introduced a "time-biasing" factor. The "modernization-factor" is a vexing problem and it is difficult to determine precisely when a criterion or reference sample becomes outdated. Because of the costs and time in developing samples, the data do not warrant a complete modernization of the inventory every ten years. After fifteen to twenty years, time biasing factors may lead to spuriously inflated results on recently developed scales if scale construction uses an earlier aggregated in-general sample. Also, after twenty years, explaining to a client that he or she has interests similar to people that were tested twenty years ago, will raise an eyebrow as to relevancy. Thus, after 10-13 years, the test publisher probably should consider seriously starting to develop the necessary plans and developmental efforts for a major revision so it could be completed within the ensuing 5 years.

Changes in Male-Female Differences

The preceding data showed that even though occupational characteristics are fairly stable over long time spans, there were also individual items that did show significant changes in popularity. The next question is: Are items that differentiate the sexes also changing over time?

If differences between male and female vocational preferences are diminishing, then at some future time an interest inventory could be developed where differential base rates of responding would not be a factor and sex biasing in test development would become a moot point. However, if differences are not disappearing, then differential

response rates by gender will have to be always accommodated into the developmental system of interest measurement.

Data presented by Campbell (1974) specifically addresses the question of changing male-female differences within occupations over 30 years. Nine occupations were available with adequate samples of males and females who had been tested during the 1930s and the 1960s. Table 7 reports the percentage responding "Like" to four items that typically show male-female differences. The first two items, "Decorate a room with flowers" and "Regular hours for work," usually are favored by females and the last two, "Repair electrical wiring" and "Express judgments openly regardless of what people say," are favored more by males.

Insert Table 7 About Here

Inspection of the data in Table 7 clearly indicates that there has not been a decrease in the magnitude of the differences between males and females within occupations from 1930 to 1968. Of the 36 contrasts (nine occupations times four items), 11 showed smaller male-female differences over time while 25 showed larger differences; differentiation in male-female responding appears to have grown larger.

The above four examples of items were not atypical; interest measurement must take into account the differences in ways that males and females respond. Until evidence is obtained that show dramatic shifts are currently occurring, waiting for these differences to disappear would be futile.

Table 7

Male-Female Differences: 1930 Versus 1968 Samples

(Table entries are percent of each sample responding "Like")

Occupations		Items											
		Decorate a room with flowers			Regular hours for work			Repair electrical wiring			Express judgements openly		
		M	F	Difference	M	F	Difference	M	F	Difference	M	F	Difference
Artists	1930	42	87	-45	39	63	-24	29	17	+12	38	32	+ 6
	1968	31	80	-51	29	48	-19	24	11	+13	60	40	+20
Lawyers	1930	21	79	-58	57	59	- 2	31	23	+ 8	55	49	+ 6
	1968	9	63	-54	30	40	-10	32	8	+24	66	49	+17
Life Insurance Sales	1930	34	85	-51	49	56	- 7	35	15	+20	41	32	+ 9
	1968	14	76	-62	14	26	-12	31	8	+23	56	38	+18
Math-Science Teachers	1930	29	82	-53	76	89	-13	55	34	+21	29	15	+14
	1968	11	70	-59	45	80	-35	48	19	+29	49	20	+29
Physicians	1930	26	80	-54	60	67	- 7	51	40	+ 9	26	28	- 2
	1968	13	71	-58	40	50	-10	53	28	+25	58	27	+31
Psychologists	1930	15	80	-65	37	60	-23	50	33	+17	34	26	+ 8
	1968	13	72	-59	27	32	- 5	38	17	+21	60	49	+11

Table 7 continued

Reporters	1930	30	83	-53	47	72	-25	22	12	+10	56	38	+18
	1968	12	78	-66	24	36	-12	25	9	+16	68	49	+19
Social Science Teachers	1930	19	90	-71	72	83	-11	36	14	+22	38	31	+ 7
	1968	7	69	-62	42	68	-26	26	7	+19	66	40	+26
YMCA-YWCA Staff	1930	31	92	-61	55	60	- 5	50	23	+27	37	21	+16
	1968	10	70	-60	36	33	+ 3	40	20	+20	35	38	- 3

SUMMARY

The scope of this paper has been to review the two major interest inventories and to explore the complexities of technical aspects in development of item sampling, norming, scoring, and reporting of results in relation to the differential treatment of sexes. Using the operational definition of sex bias as any factor that may influence a person or others to limit career opportunities solely on the basis of gender, the following areas were investigated and recommendations made for avoiding potential sex bias.

Item Sampling

Interest inventories should not have separate forms of the instrument for males and females. Potential sex biasing should be eliminated at the item development level by using the same set of items for both males and females. Special care has to be excercised so that an item does not inherently make it more applicable to either gender. For example, items such as "Policeman" and "Policewoman" are less desireable than "Police Officer," or "Realtor" would be preferable to "Real Estate Salesman" or "Real Estate Salesperson." For some types of items, reference to both genders may be the only viable alternative, such as "Dressmaker/Tailor" or "Waiter/Waitress." A recommended procedure for the test developers would be to request the U.S.-GSA Civil Rights Office to review item content of new inventories and suggest changes for any items that appear to be discriminating to minority groups.

Scale Development

There are various factors that have to be considered within the context of sex bias when scales are constructed for interest inventories.

There are two main types of scales (externally and internally based scales) constructed for interest inventories and each has its own distinctive issues relating to sex bias.

Externally-based Scales

Externally-based scales (occupational or criterion scales) are based on item responses of appropriate criterion samples. For example, samples of employed artists, mathematicians, and elementary teachers are collected and their responses to the inventory are tabulated. A person's score on criterion scales measures the extent of similarity with those in a criterion sample. Of immediate concern in the development of criterion scales is the extent and magnitude of differences between male and female responses and how to accommodate these differences. For example, should males and females be combined in the criterion sample? Should scores be reported separately for males and females? Is the magnitude of male-female differences diminishing with growing awareness that traditional sex roles are arbitrary? Should scores be reported on scales developed on opposite-sex criterion groups when appropriate-sex scales are not available? Do the criterion samples become out-dated and need periodic revision?

Empirical evidence is very consistent in showing that males respond differently than females to many items on interest inventories. These differences are neither small nor infrequent. Differential responses of males and females should not be ignored; combining responses of males and females in criterion samples would decrease the predictive and concurrent validity of empirical criterion scales. Furthermore, the magnitude of differences between males and females has not substantially diminished during the past 30 years.

Since these male-female differences appear to be fairly stable and large, occupational scales should be developed on male criterion samples and female criterion samples. Every effort should be made to collect adequate samples of males and females for each occupation represented on the inventory's reporting of results. Furthermore, in development of empirical scales for both sexes, caution should be exercised so as to not differentially include items that represent non-valid male-female differences.

When an appropriate sex criterion sample is not available, and so as to not limit career options, scores should be reported for scales based on the opposite sex criterion sample but the user should be cogently aware of the impact that sexual stereotypes will have on the resultant score.

Stability of response preferences also are evidenced for occupational samples. Occupational scales developed in the past still show excellent concurrent validity 30 years later. However, when new scales are developed for an old inventory, a "modernization" factor occurs in the content of the new scales. Because of cost and time factors in revising an inventory, empirical data do not warrant a recommended nor necessary revision every 10 years. If newly developed scales are added to existing sales on an inventory developed 15-20 years earlier, this modernization factor is likely to occur. Thus interest inventories should be revised at least after 20 years and preferably after 15 years. This revision would update the content of items and develop occupational scales that are more responsive to future changes in occupational patterns in existence within society.

Internally-based Scales

Internally-based scales (homogeneous or basic interest scales) are based on interrelationships of items within the inventory. These scales measure content areas of interests such as Mechanical interests, Art interests, Social Service interests and are more relevant to the measurement of avocational types of interests. Because of the nature of scale development (item intercorrelations) and the relatively lesser importance of predictive validity than for criterion scales, responses for males and females could be combined into a general sample for computing the necessary statistics for construction of the scales. Using one sample with equal representation of males and females would lessen the impact of potential sex biasing, but differential response frequencies of males and females would still have to be a concern. The most meaningful solution would be to use the combined general reference sample as the norm sample for converting raw scores to standardized scores but providing for separate interpretive norm distributions for males and females. Thus, scores would be equivalent but would have different interpretive meanings dependent upon gender.

Composition of an appropriate norm reference group for internally-based scales would depend on the intended usage of the inventory and judgment of the test developer. For example, if the inventory is intended primarily for the armed services, the norm sample should be composed of armed service personnel and not junior high school students; or, the test developer may deem that a general sample of adult males and females would be a more appropriate reference sample than college graduates. In any event, reference groups could be composed of equal representation

of males and females if results are presented with separate interpretive band widths that indicate male-female differences.

This use of a single norm sample also permits an easy linkage to scores of male and female criterion groups on these homogeneous scales since the standard scores would be equivalent in meaning. Updating of norm reference samples should be as frequent as that discussed previously for updating of occupational samples.

Socialization Bias

Many of the technical difficulties outlined above are a result of the substantial sex differences to many interest inventory items. The possibility of developing an effective inventory free of such bias seems remote. The SVIB has undergone numerous studies over 40 years to increase its validity, and major revisions in the last 10 years have eliminated less valid items, archaic items, and items that inherently favor one gender. The result will be a single inventory to be released in 1974 where considerable attention has been given to sex biasing factors. Still over one-half of the items show significant male-female differences. Eliminating items that are not free of socialization bias would seriously decrease the validity and viability of the instrument. Items such as "Artist" and "Decorating a room with flowers" show considerable male-female differences but are very valid types of items in differentiating interest of interior decorators from other occupations. Data indicate that changes within society will have to be more dramatic than has occurred during the last 30 years in reducing male-female stereotypes before effective inventories free of socialization can be constructed.

Interpretive Materials

Any published interest inventory should have a sound and comprehensive test manual following the guidelines of the American Psychological Association as outlined in Standards for Educational and Psychological Tests and Manuals (1966). In addition to recommended standards of describing criterion groups, validity, reliability, and so forth, a fairly in-depth discussion should be included in the manual and other interpreitve materials elucidating the extent and nature of male-female differences on the inventory.

Test developers should outline what procedures were used to handle male-female response differences during scale construction, norming, and reporting of results for the various types of scales that the inventory may contain. Specific mention should be made as to what efforts were made to assure the applicability of the items to both genders. Data should be presented indicating the extent that scales developed on either male or female criterion groups can be generalized or used for the opposite sex.

Sufficient Time and Money

If sufficient time and money were available, what directions could be taken in the development of a new inventory that would lessen or eliminate the socialization effect? It may be possible to write a sufficient number of items that would make a viable inventory and still show no large differences between sexes. This would involve an extensive developmental effort and continual trying out of items on various groups of subjects to ascertain the instrument's validity.

Still, even though enough items may be aggregated that are relatively free of sex biases, small male-female differences may have a cumulative

effect during construction and norming of the resultant scales. If sex differences existed in the scales, extensive validity studies would have to be done to measure the extent of the impact of these differences. Also, the data may show that for subjects who have well-defined and strong career aspirations, male-female differences are a minor factor and scales could be developed so as to not have a limiting effect on either gender. If results pointed in this direction, use of career commitment may be a more meaningful moderator variable than gender.

Various statistical procedures could be tried to compensate for male-female response differences, such as multivariate discriminant analysis, but such procedures would be applying differential weights, based on gender, to results and considerable analyses would have to be done to show that these procedures do not decrease the validity of the inventory for both sexes.

In the long run, writing items that are as valid as current items and show no male-female differences may be easier than to try to statistically correct for these differences on a post hoc basis.

FOOTNOTE--

[1]The author is deeply indebted to Dr. David P. Campbell, Director of the Center for Interest Measurement Research, University of Minnesota, for much of the empirical data contained in this report.

REFERENCES

American Psychological Association. Standards for psychological and educational tests. Washington, 1966.

Berdie, R.F. Intra-individual variability and predictability. Educational and Psychological Measurement, 1961, 21, 663-676.

Campbell, D.P. The stability of vocational interests within occupations over long time spans. Personnel and Guidance Journal, 1966(a), 44, 1012-1019.

Campbell, D.P. Manual for the Strong Vocational Interest Blanks. Stanford: Stanford University Press, 1966(b).

Campbell, D.P. Changing patterns of interests within the American society. Measurement and Evaluation in Guidance, 1968, 1, 36-49.

Campbell, D.P. Handbook for the Strong Vocational Interest Blank. Stanford: Stanford University Press, 1971.

Campbell, D.P. Assessing Vocational Interests: A Manual for the Strong-Campbell Interest Inventory. Stanford: Stanford University Press, 1974 (in press).

Clark, K.E. Vocational Interests of Nonprofessional Men. Minneapolis: University of Minnesota Press, 1961.

Clark, K.E. and Campbell, D.P. Manual for the Minnesota Vocational Interest Inventory. New York: Psychological Corporation, 1965.

Cole, N.S. On measuring the vocational interests of women. Journal of Counseling Psychology, 1973, 20, 105-112.

Cole, N.S. and Hanson, G.R. An analysis of the structure of vocational interests. Journal of Counseling Psychology, 1971, 18, 478-486.

Darley, J.G. and Hagenah, T. Vocational interest measurement. Minneapolis: University of Minnesota Press, 1955.

Dunnette, M.D. and Kirchner, W.K. Validation of psychological tests in industry. Personnel Administration, 1958, 21, 20-27.

Ferguson, L.W. Life insurance interest, ability, and termination of employment. Personnel Journal, 1958, 11, 189-193.

Ferguson, L.W. Ability, interest, and aptitude. Journal of Applied Psychology, 1960, 44, 126-131.

Frederiksen, N. and Gilbert, A.C. Replication of a study of differential predictability. Eudcational and Psychological Measurement, 1960, 20, 759-767.

Frederiksen, N. and Melville, S.D. Differential predictability in the use of test scores. Educational and Psychological Measurement, 1954, 14, 647-656.

Ghiselli, E.E. Moderating effects and differential reliability and validity. Journal of Applied Psychology, 1963, 47, 81-86.

Gough, H.G. Manual for the California Psychological Inventory. Palo Alto: Consulting Psychologists Press, 1969.

Grooms, R.R. and Endler, N.S. The effect of anxiety on academic achievement. Journal of Educational Psychology, 1960, 51, 229-304.

Hornaday, J.A. and Kuder, G.F. A study of male occupational scales applied to women. Educational and Psychological Measurement, 1961, 21 859-864.

Huth, C.M. Measuring women's interests: how useful? Personnel and Guidance Journal, 1973, 51, 539-545.

Johnson, R.W. and Johansson, C.B. Moderating effects of basic interests on predictive validity of SVIB occupational scales. Proceedings of the 80th Annual Convention, American Psychological Association, 1972, 7, 589-590.

Johansson, C.B. Psychometric characteristics of the inventory of interests. Unpublished Ph.D. dissertaion, University of Minnesota, 1969.

Johansson, C.B. The Strong Vocational Interest Blank in-general groups. Journal of Counseling Psychology, 1974 (in press).

Johansson, C.B. and Campbell, D.P. Stability of the Strong Vocational Interest Blank for men. Journal of Applied Psychology, 1971, 55, 34-36.

Johansson, C.B. and Harmon, L.W. The Strong Vocational Interest Blank: one form or two? Journal of Counseling Psychology, 1972, 19, 404-410.

Kuder, G.F. Manual for the Occupational Interest Survey. Chicago: Science Research Associates, 1966.

Kuder, G.F. Manual for Kuder General Interest Survey-Form E. Chicago: Science Research Associates, 1973.

Kuder, G.F. Kuder DD Occupational Interest Survey Interpretive Leaflet. Chicago: Science Research Associates, 1974.

Laime, B.F. and Zytowski, D.G. Women's scores on the M and F forms of the SVIB. Vocational Guidance Quarterly, 1964, 12, 116-118.

Matarazgo, J.D., Allen, B.V., Saslow, G., and Wiens, A. Characteristics of successful policemen and firemen applicants. Journal of Applied Psychology, 1964, 48, 123-133.

Schlossberg, N.K. and Goodman, J. Imperative for change: counselor use of the Strong Vocational Interest Blank. Impact, 1972, 2, 25-29.

Seashore, H.G. Women are more predictable than men. Presidential address, Division 17, American Psychological Association Annual Convention, New York, September, 1961.

Strong, E.K., Jr. Vocational interests of men and women. Stanford: Stanford University Press, 1943.

Strong, E.K., Jr. Manual for the SVIB. Pa o Alto: Consulting Psychologists Press, 1959.

Super, D.C. and Crites, J.O. Appraising vocational fitness. New York: Harper and Row, 1962.

Thrush, R.S. and King, P.T. Differential interests among medical students. Vocational Guidance Quarterly, Winter, 1965, 120-123.

Welsh, G.S. and Dahlstrom, W.G. Basic readings on the MMPI in psychology and medicine. Minneapolis: University of Minnesota Press, 1956.